

Finding new candidate genes associated with molecular systems by computationally screening for typical phenotypes.

van Hoek, I.N. and Huijnen, M.A.

10/10/2018

Finding new candidate genes associated with molecular systems by computational screening for typical phenotypes | 2018

Finding new candidate genes associated with molecular systems by computational screening for typical phenotypes.

Abstract

The observance of phenotypes can be of important value in unraveling mechanisms of genetic disorders. Fortunately, human phenotypes linked to genetic defects are documented in online databases, one of which was built for facilitating computational analysis, the Human Phenotype Ontology (HPO) database. The observation of phenotypes on such a large scale has the potential of uncovering unknown molecular pathways. With mitochondrial disorders as research focus, mitochondrial proteins served as input in our new phenotype analysis tool, Galiphy. HPO phenotypes were scored based on their prevalence among the mitochondrial proteins prior to the calculation of the gene scores. All 3,246 genes in HPO were ranked on their scores after which statistical validation demonstrated that Galiphy could rank the genes according to their class, corresponding to areas under the ROC curves of $\geq 95\%$. Therefore, we propose to have successfully generated a 'phenotype fingerprint' using a simple but elegant scoring method. Specifically with this input, it revealed candidate genes which have a possible role in mitochondrial dysfunction. We conclude that ranking phenotypes using our bioinformatics tool can play a meaningful role in the set-up of experimental searches towards unravelling the molecular mechanisms behind genetic defects. Galiphy is freely available at <https://www3.cmbi.umcn.nl/galiphy>.

Table of Content

| | |
|----------------------|----|
| Abstract | 3 |
| Introduction | 5 |
| Material & Methods | 7 |
| Results & Discussion | 8 |
| References | 12 |
| Appendices | 13 |

Ilja N. van Hoek

Radboud University Nijmegen

M.A. Huijnen

Centre for Molecular and Biomolecular Informatics;

Radboud Centre for Mitochondrial Medicine;

Radboudumc, Nijmegen, The Netherlands

INTRODUCTION

Ever since mutations and phenotypes were considered linked, phenotypes have played a key role in genetic research. Among the first scientists to use this link were Beadle and Tatum in the 1940s¹ with their revolutionary pathway experiment. They deduced the arginine biosynthesis pathway from phenotype analysis of mutated *Neurospora*.

We used a similar deduction on a large-scale computational analysis, as we compiled a 'phenotype fingerprint' of a gene set by using phenotype data. In our approach, we used phenotype data from the Human Phenotype Ontology (HPO) database², which is the most comprehensive database with phenotypes linked to genes. HPO converted phenotypes from the Online Mendelian Inheritance in Man (OMIM) resource and the biomedical literature into a structured ontology of phenotype terms. This ontology includes over 11,000 phenotype terms, annotated to 3,246 disease genes. Our algorithm weighs the HPO phenotype terms according to their prevalence in a query gene set, as

demonstrated in the grey box in Figure 1. Subsequently, all phenotypes are mapped back to their genes, scored and ranked by their associated weighted phenotypes (see right side of Figure 1). If the query genes share a functional relationship, the ranking can support the prediction of new genes.

In the study described in this report, we aim to shed light on the mechanisms of secondary mitochondrial disorders. The proteins affected in secondary mitochondrial disorders are not known to be associated with mitochondria, unlike primary mitochondrial disorders. Mitochondria are essential for many metabolic and cellular processes in the eukaryotic cell, including energy metabolism. Dysfunction of the mitochondria can result in disorders affecting the brain, skeletal muscle, heart, and liver. However, the clinical features are rarely pathognomonic³, characterizing one of the main challenges encountered in diagnosing mitochondrial disorders (MD). This is because the exact same list of phenotypes can be caused by different mutations and,

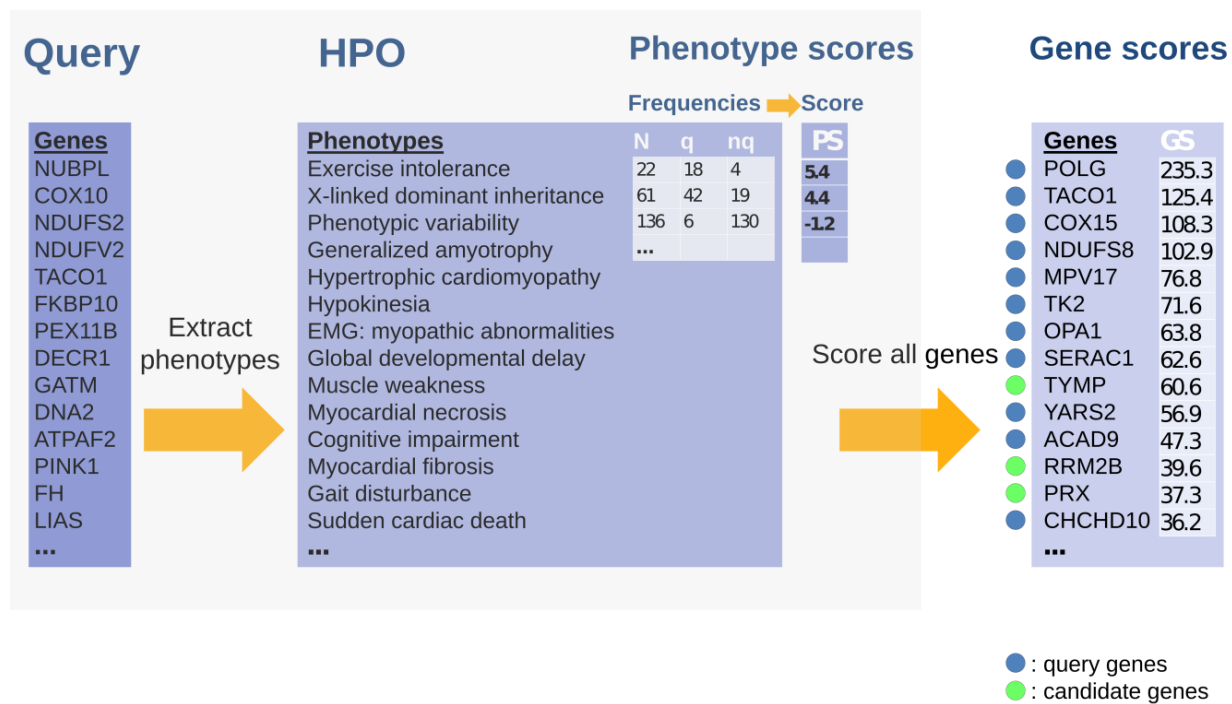


Figure 1. Schematic overview of Galiphy. The goal of the tool is to find new genes that have a similar phenotype as the query gene set. Here, mitochondrial proteins are used as query gene set. First, from a query gene set, HPO associates phenotypes to e.g. 80 of the 300 query genes, of which the phenotypes are extracted. Second, the abundance of each phenotype among the gene sets is determined ('q' are query genes, whereas 'nq' are non-query genes and 'N' are all genes in HPO) after which the phenotype score ('PS') can be calculated (see Equation 1). Lastly, for each gene present in HPO a gene score ('GS') can be determined by the combining of all phenotype scores (see Equation 2). Genes outside the initial query set are depicted with a green dot.

vice versa, same mutations can lead to different phenotypes.⁴ For this reason, whole exome sequencing (WES) has recently been implemented as a state-of-the-art diagnostic test for MD by Wortmann *et al.*⁴ In their study, histological, clinical, neuroradiological, metabolic, and biochemical data of 109 undiagnosed patients were thoroughly evaluated, after which they divided the patients into three groups based on level of suspicion of having MD. Subsequently, the WES analysis revealed 21 mutations in genes which were hitherto unknown to be localized to the mitochondrion, of which seven different genes were mutated in the patients with high suspicion of MD (SLC3A1/PREPL, NGLY1, ARID1B, SCN1A and three unpublished findings).

This leads us to our research question: how can a non-mitochondrial protein be responsible for mitochondrial dysfunction? Of which for instance two theories can be hypothesized. 1. Although Mitocarta⁵ is the most comprehensive inventory of proteins localized in the mitochondrion, proteins absent from this list could be incorrectly designated as non-mitochondrial. 2. Alternatively, a non-mitochondrial protein could affect a pathway, which in turn influences mitochondrial function. In both cases, an unknown relationship with mitochondrial function is to be discovered. We therefore reasoned that Galiphy could help investigating the mechanisms of mitochondrial dysfunction by providing prioritization of genes based on mitochondrial typical phenotypes.

MATERIAL & METHODS

Data sets

To collect phenotypes caused by disease genes, a phenotype database was required that contained appropriate annotation of phenotypes to genes. A controlled vocabulary of phenotypes structured in a manually curated systematic ontology was provided by the Human Phenotype Ontology (HPO) database². They specifically developed a human phenotype database to facilitate comparison of phenotypes. HPO terms are organized in an ontology, in which terms are arranged in “is-a” relationships. This relationship is transitive, meaning that a genes that is annotated to a child term (specific HPO term) inherits all parent terms (more general HPO terms) up all paths to the root.

The file containing the *genes-to-phenotypes* annotations was downloaded (on 11/26/2015, version 5.1.73), restricting ourselves to the most specific HPO terms for the genes. This file links 5,881 different phenotype terms to 3,246 genes, which annotates 26 phenotype terms per gene on average.

In order to collect all human mitochondrial proteins for our query gene lists, the Human Mitocarta⁵ version 2.0 was downloaded on 11/26/2015.

Score calculation procedure

As a query for the algorithm, a set of genes was established. Subsequently, we defined a formula for phenotypes based on the following principles: 1. Phenotypes that are more often associated to the query genes should score higher than phenotypes that are more often associated to the non-query genes. 2. If, for instance, HPO annotates 300 genes of the query, the remaining 3000 genes are automatically defined as non-query genes. 3. Prior is taken into account, meaning that the relative chances of finding phenotypes in each query list are incorporated. For each phenotype, we define its *Phenotype Score* (PS) as:

$$PS = \log_2 \frac{q|Q}{nq|nQ} \frac{total\ Q}{total\ nQ}$$

Equation 1

where q is the abundance of the phenotype among the query gene list Q , and nq is the abundance of the phenotype among the non-query gene list nQ . Note, if a phenotype is associated with zero genes in a gene set, its frequency parameter (n or nq) is set to 0.1 to avoid dividing by zero. Once all 5881 phenotypes were given a score, the phenotypes were mapped back to their associated genes. For each gene, its phenotype set is

defined by $P = \{p_1, p_2, \dots, p_i\}$, and a *Gene Score* ‘GS’ was calculated by:

$$GS = \sum_{p_i \in P} PS$$

Equation 2

This score represents to which extent the gene is associated to phenotypes typical for the query gene set. After all HPO genes were given a gene score, further analysis was performed to evaluate the performance of the algorithm.

Query gene lists

The first query gene list was compiled by selecting all Mitocarta genes that were annotated by HPO, remaining 351 genes. Analysis of the first output revealed that six of the 351 mitochondrial proteins caused a bias in the algorithm. The phenotypes of the succinate dehydrogenase genes were not congenital, these genes were therefore eliminated from the second query gene list. To make the scoring specific for MD phenotypes, the mitochondrial protein list was manually curated and 168 proteins directly responsible for energy conversion were selected for the third query gene list. The genes in the three different queries are listed in Appendix I.

Performance analysis of Galiphy

We conducted a gene set enrichment analysis (GSEA⁶) to test the robustness of the algorithm. The null hypothesis of GSEA is defined as follows: genes belonging to predefined classes are randomly distributed when ranked according to a specific parameter. Thus by ranking all HPO genes by their gene score, the quality of the division between query genes and non-query genes was measured. In order to perform Bayesian analyses, we define the negative gene set as the non-mitochondrial proteins that were in the Mitocarta training set of Calvo *et al*⁶, classified as ‘non_mito’. Positive predictive value graphs and receiving operating characteristic curves were made for each query gene list to illustrate how the query genes are represented among the highest scoring genes.

Further functional analyses were performed on the high scoring non-query genes (i.e., candidate genes) in order to gain insight in the possible relation with MD. String⁷ version 10.0 was used for the top 200 candidate genes (on 6/14/2016). For genes that had individually caught our attention, we obtained co-expression information from the WeGET⁸ tool (data obtained on 4/6/2016, version 1.0).

RESULTS & DISCUSSION

Using a query of mitochondrial genes, our algorithm generated an output of specificity scores for all phenotypes and genes in the HPO database. The query gene list consists of the genes linked to at least one phenotype in the HPO database. Phenotypes were scored according to their prevalence among query genes and non-query genes. These phenotype scores are representative for the level of particularity of the phenotype for the query genes. It is important to note that observation of the phenotype scores is not only relevant for fundamental biochemical questions, but also for clinical applications. The phenotype scores (PS) combined resulted in a gene score (GS) (see Equation 1 and 2 in *Material & Methods*). The establishment of the GS of two example genes are demonstrated in Figure 2. The scoring of a gene indicated how typical its phenotypes are for the query gene set. When the list of all genes were rank-ordered by GS, the high scoring non-query genes, i.e., candidate genes were further analyzed. In this study, the high scoring non-query

genes were of particular interest and therefore studied in more detail. This could enlighten processes linked to the query genes. In addition, the candidate genes were further analyzed with the String enrichment tool.

Establishing adjusted query gene sets

The first query generated an output of which the top 50 highest GS (ranging from 281.91 to 86.91) contained five complex IV genes, five complex II and 23 complex I genes. The top 50 highest scoring phenotypes were examined by a physician specialized in MD. Surprisingly, some of these phenotypes were assigned as not typical for MD. The succinate dehydrogenase genes caused the appearance of tumor-related phenotypes to score among the highest ranked phenotypes (as is illustrated in Appendix II). Although its mechanism is not fully understood, dysfunction of the complex II succinate dehydrogenase can lead to the formation of pheochromocytoma (PHEO) and paraganglioma (PGL)⁹. The genes caused bias in the algorithm, as genes that

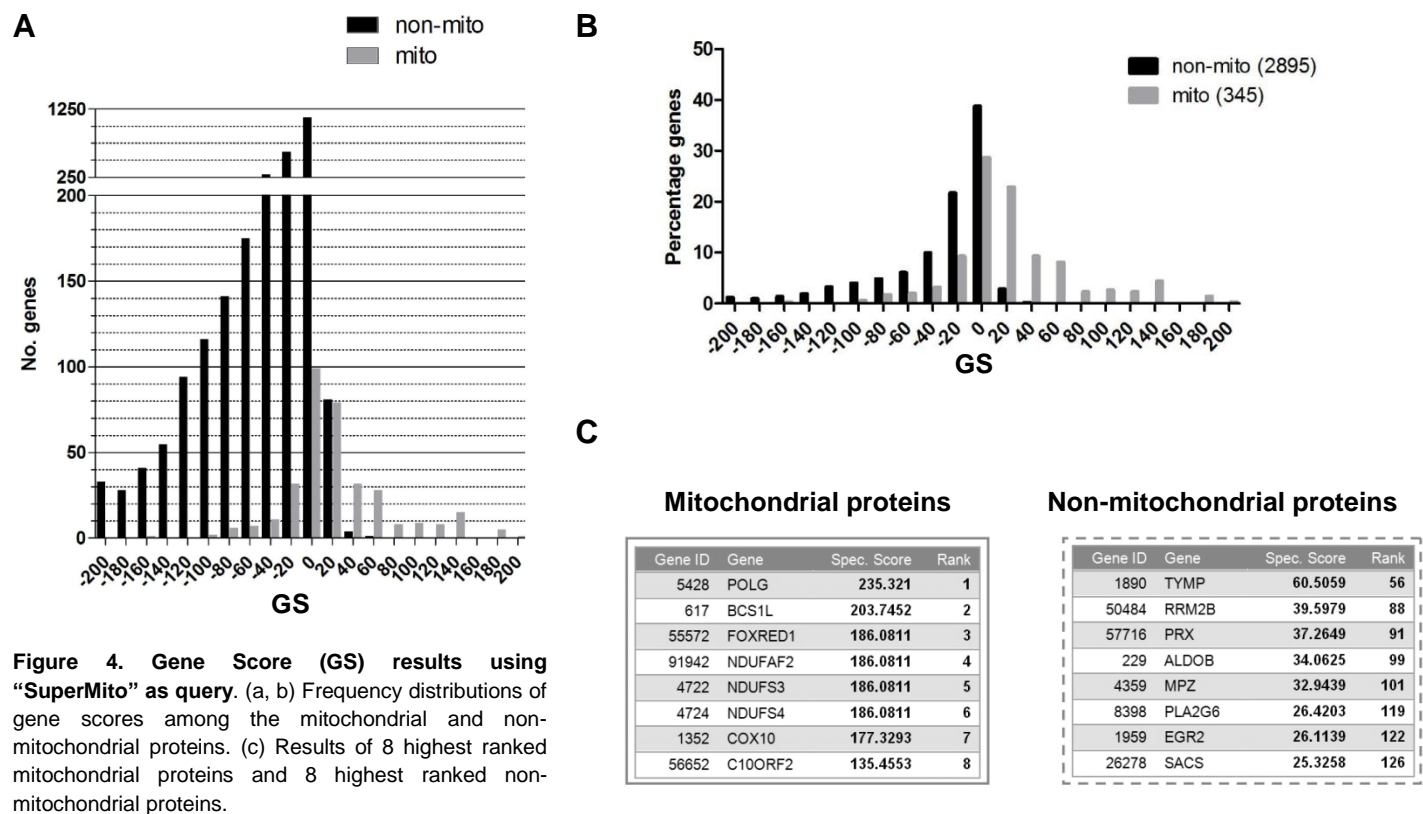
Figure 2. Phenotype scoring and subsequent gene scoring of two example genes. HPO assigned 21 phenotypes to CD40LG. The phenotype "Hepatomegaly" is assigned to 339 (N) out of 3246 genes, of which 29 (q) out of 168 genes are query genes and 310 (nq) out of 3079 are non-query genes. Each phenotype is scored as shown in the last column. This phenotype score represents how specific the phenotype is for the query genes. A higher phenotype score indicates that the phenotype is 'typical' for the query genes. The phenotype scores of all 21 phenotypes linked to CD40LG are combined and the resulting gene score is shown in the last row. A higher gene score indicates that the gene has a phenotypically outlook which is similar to the query genes.

A

| Gene ID | Gene name | Phenotype ID | Phenotype description | N | q | nq | Phenotype score |
|---------------------------|-----------|--------------|----------------------------------------|-----|-----|-----|-----------------|
| 959 | CD40LG | HP:0002240 | Hepatomegaly | 339 | 29 | 310 | 0.78 |
| 959 | CD40LG | HP:0005479 | IgE deficiency | 2 | 0.1 | 2 | -0.13 |
| 959 | CD40LG | HP:0001873 | Thrombocytopenia | 158 | 6 | 152 | -0.47 |
| 959 | CD40LG | HP:0002847 | Impaired memory B-cell generation | 3 | 0.1 | 3 | -0.71 |
| 959 | CD40LG | HP:0002849 | Absence of lymph node germinal center | 3 | 0.1 | 3 | -0.71 |
| 959 | CD40LG | HP:0002961 | Dysgammaglobulinemia | 3 | 0.1 | 3 | -0.71 |
| 959 | CD40LG | HP:0010280 | Stomatitis | 3 | 0.1 | 3 | -0.71 |
| 959 | CD40LG | HP:0001419 | X-linked recessive inheritance | 109 | 3 | 106 | -0.95 |
| 959 | CD40LG | HP:0002014 | Diarrhea | 116 | 3 | 113 | -1.04 |
| 959 | CD40LG | HP:0002959 | Impaired Ig class switch recombination | 4 | 0.1 | 4 | -1.13 |
| 959 | CD40LG | HP:0005419 | Decreased T cell activation | 4 | 0.1 | 4 | -1.13 |
| 959 | CD40LG | HP:0001875 | Neutropenia | 58 | 1 | 57 | -1.64 |
| 959 | CD40LG | HP:0012115 | Hepatitis | 7 | 0.1 | 7 | -1.93 |
| 959 | CD40LG | HP:0003496 | Increased IgM level | 8 | 0.1 | 8 | -2.13 |
| 959 | CD40LG | HP:0001744 | Splenomegaly | 241 | 2 | 239 | -2.71 |
| 959 | CD40LG | HP:0002720 | IgA deficiency | 17 | 0.1 | 17 | -3.21 |
| 959 | CD40LG | HP:0000230 | Gingivitis | 23 | 0.1 | 23 | -3.65 |
| 959 | CD40LG | HP:0004315 | IgG deficiency | 23 | 0.1 | 23 | -3.65 |
| 959 | CD40LG | HP:0002718 | Recurrent bacterial infections | 49 | 0.1 | 49 | -4.74 |
| 959 | CD40LG | HP:0001878 | Hemolytic anemia | 56 | 0.1 | 56 | -4.93 |
| 959 | CD40LG | HP:0002721 | Immunodeficiency | 61 | 0.1 | 61 | -5.06 |
| CD 40LG gene score | | | | | | | -40.55 |

B

| Gene ID | Gene name | Phenotype ID | Phenotype description | N | q | nq | Phenotype score |
|------------------------|-----------|--------------|-------------------------------------|------|-----|------|-----------------|
| 966 | CD59 | HP:0003690 | Limb muscle weakness | 24 | 6 | 18 | 2.61 |
| 966 | CD59 | HP:0003202 | Skeletal muscle atrophy | 220 | 36 | 184 | 1.84 |
| 966 | CD59 | HP:0001252 | Muscular hypotonia | 701 | 95 | 606 | 1.52 |
| 966 | CD59 | HP:0001284 | Areflexia | 115 | 13 | 102 | 1.22 |
| 966 | CD59 | HP:0000007 | Autosomal recessive inheritance | 1748 | 129 | 1619 | 0.55 |
| 966 | CD59 | HP:0002922 | Increased CSF protein | 18 | 1 | 17 | 0.11 |
| 966 | CD59 | HP:0004818 | Paroxysmal nocturnal hemoglobinuria | 3 | 0.1 | 3 | -0.71 |
| 966 | CD59 | HP:0001878 | Hemolytic anemia | 56 | 0.1 | 56 | -4.93 |
| CD59 gene score | | | | | | | 2.21 |



cause cancerous phenotypes are not congenital defects. In order to keep the query a gene list with only congenital genes, scores were generated with a second query in which the six SDH genes were excluded. The output of this query “MitoSDHrem” list was similar to the first output: for example, among the top 50 highest GS (now ranging from 186.76 to 68.97) the same complex IV and complex I genes appeared. The genetic defects leading to mitochondrial disorders were of main interest in this research. Therefore, we changed our focus on genes in which mutations cause MD. We hypothesized that restricting the input to only 168 genes responsible for oxidative phosphorylation function would result in an output which prioritizes more specific for mitochondrial disorders than the previous two queries. Using this “SuperMito” query indeed resulted in specific scoring phenotypes for congenital MD (Appendix III). The gene scores (Figure 3a and b) indicated how typical the gene is associated with the query genes’ phenotypes. A distribution of the phenotype scores for the different query genes can be found in Appendix IV. For simplicity, the following analysis of the prioritized genes only describe the results from the “SuperMito” query.

Candidate genes

Interestingly, some of the candidate genes have previously been suggested to be associated with mitochondrial function (Figure 3c). Mutations in both

TYMP and RRM2B are known to cause mitochondrial DNA depletion syndromes.¹⁰ The PLA2G6 gene has been suggested to be involved in mitochondrial function.¹¹ We used the top 200 highest scoring non-mitochondrial proteins (i.e., our candidate genes) for analysis with String⁷, which placed the genes in a context of an association network. In addition, String provided us the enrichment of the genes among KEGG pathways and GO biological processes. The genes were highly represented in metabolic processes and pathways, as well as N-glycan biosynthesis, bile secretion and the lysosome pathway (Appendix III). We sought additional support for linking the candidate genes to mitochondria by including co-expression information. We used a p-value quantification provided by the weighted gene co-expression tool WeGET⁸, using the “SuperMito” list as a query.

Statistical evaluation

In addition to the manual analysis of the output, the performance of the algorithm was also evaluated statistically. To illustrate the ability of the scoring method to classify mitochondrial genes, all HPO genes were rank-ordered by their score and gene set enrichment analysis⁶ (GSEA) was performed (Figure 4a and b). In addition, receiver operating characteristic (ROC) analysis allowed evaluation of the predictive classification performance of the algorithm (Figure 4c).

Lastly, the precision of the algorithm was demonstrated by a precision rank plot (Figure 4d). No cross-validation was used as statistical measurement, as the scoring method did not allow setting a threshold.

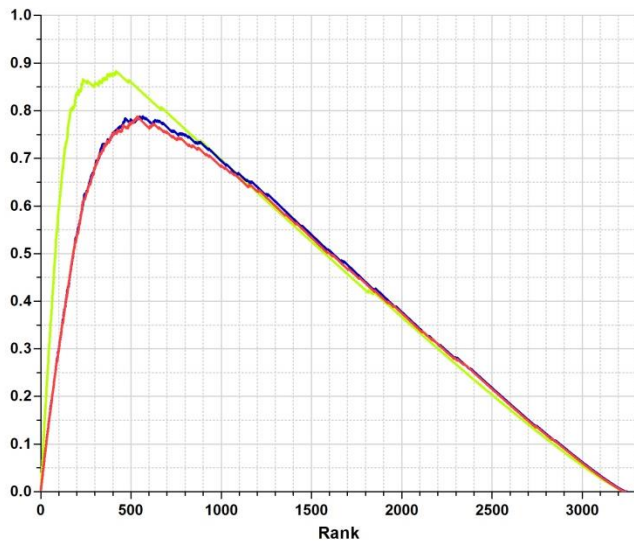
Considering that the “MitoAll” and “MitoSDHrem” had only a difference of six genes so their overall similar performance was not surprising. Fortunately, the overall

higher performance of the “SuperMito” gene set satisfied our expectations.

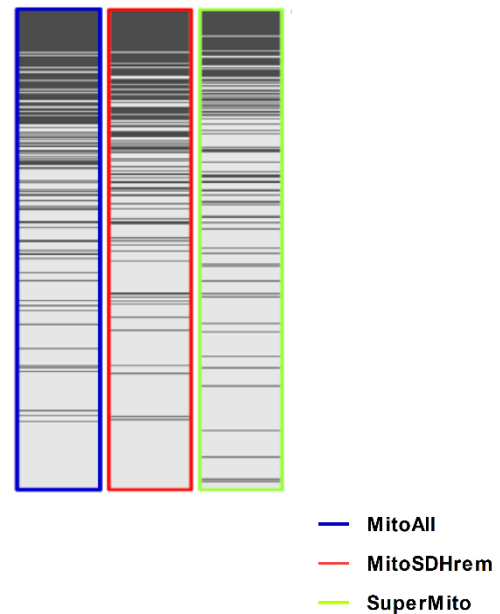
Selection of a gene for experimental validation

Twelve genes from the WES analysis list of patients with secondary mitochondrial disorders mentioned earlier were among the top 300 candidate genes. Also, three

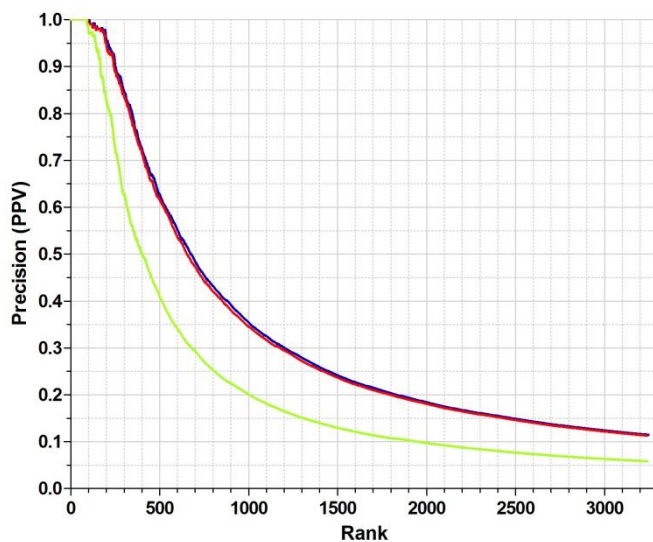
A



B



C



D

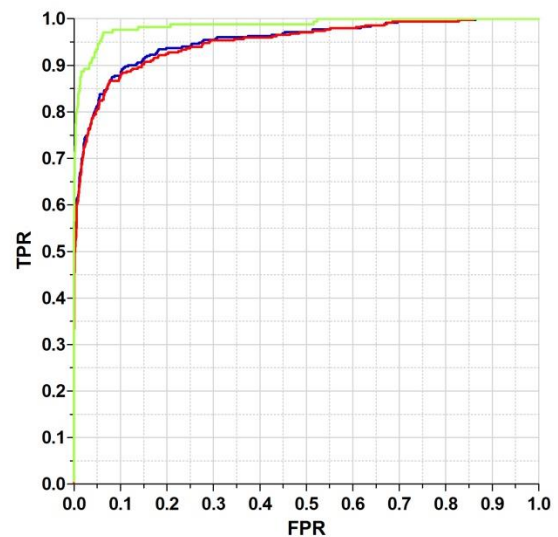


Figure 5. Statistical analyses depict that genes are ranked according to their mitochondrial phenotype. (a) Gene set enrichment plot shows that query genes are overrepresented among top-ranked genes, i.e. the leading edge set of genes before the peak of the graph. The peaks are found at rank 417, 535 and 564 for the SuperMito gene set (green), the MitoSDHrem gene set (red) and the MitoAll gene set (blue), respectively. (b) Schematic representation of disposition of the query genes in the rank-sorted list of all 3246 genes. (c) Precision rank plot shows that 50% of the genes are mitochondrial at rank 401, 653 and 674 for the gene sets SuperMito, MitoSDHrem and MitoAll, respectively. (d) Receiving operating characteristic graph tests the predictive ability of the scoring for the three query gene lists. AUC of the gene sets are 0.957, 0.949 and 0.952, for SuperMito, MitoSDHrem and MitoAll, respectively.

genes were significantly co-expressed with the “SuperMito” genes (Appendix VI). The NGLY1 gene and the EIF2B3 gene both belonged to these two groups. For NGLY1 mitochondrial association was suggested in literature¹².

Considerations of the tool

We introduced a new phenotype tool which is able to rank the genes associated to mitochondrial disorders according to their weighed phenotypes. Our results show that this type of scoring can be used in exploring gene functions. Moreover, Galiphy is also potentially useful in clinical applications. In essence, we only used simple computational methods, which is clearly an advantage as it can be exploited by every researcher or physician.

In the meantime, we will seek to improve our tool, which can be done in the following ways.

We encountered an issue regarding the ontology tree in the phenotype database. The HPO database assigns phenotypes to genes, but some phenotypes are not described at detailed level, because for example the heterogeneity of a genetic defect or general symptom of a disorder. In contrast, a specific phenotype could be caused by a thoroughly studied disorder, particular symptoms of the disorder or a regular genetic defect. As a result, difference in detail level of phenotype leads to problems in the scoring method. The phenotype ‘Nausea and vomiting’ illustrates this ontology issue: this phenotype is linked to 79 genes, its child term ‘Nausea’ and ‘Vomiting’ are linked to 11 and 106 genes, respectively. These inconsistent frequencies cause propensity in the phenotype scoring because the phenotypes are not correlated as one would expect. Further research could be done on this ontology, where it must be investigated whether the use of cumulative frequency up all paths to the root could result in a better scoring output.

Another suggestion for research is to find whether there are correlations when observing the phenotypes of each gene, and their ontology parent terms. So there might be some phenotypes overrepresented originating from the same parent term in the ontology tree. This way, it can be established whether genes have phenotypes annotated which share a more general phenotype. If possible and pertinent, the frequencies of the shared phenotype can be combined within the scoring of one gene.

Other tools that use phenotype information for exploration of gene function, include PHIVE¹³, eXtasy¹⁴,

Phen-Gen¹⁵ (side project of HPO), Phevor¹⁶. However, all these tools require genomic information about patients of a disorder, i.e. SNVs (eXtasy), exome variance (PHIVE) and sequence data (Phen-Gen and Phevor). Whereas Galiphy only requires a simple list of genes of interest.

Overall, Galiphy differs from other phenotype exploration tools, mainly due to its simplicity: a set of multiple genes is used as a query and phenotypic data is used for making a ‘phenotype fingerprint’. The researcher or physician solely needs an interest in a set of related genes, for instance a common disorder or a common pathway.

Availability and implementation

Galiphy is available under the following link: <https://www3.cmbi.umcn.nl/galiphy>. The scripts are available at <https://github.com/iljavh/galiphy>.

Contact

I.N. van Hoek; iljavanhoek2@gmail.com

REFERENCES

1. Beadle, G.W. & Tatum, E.L. (1941). Genetic Control of Biochemical Reactions in *Neurospora*. *Proceedings of the National Academy of Sciences of the United States of America*, 27, 499-506.
2. Groza, T., Köhler, S., Moldenhauer, D., Vasilevsky, N., Baynam, G. Zemojtel, T., ... Robinson, P.N. (2015). The Human Phenotype Ontology: Semantic Unification of Common and Rare Disease. *American Journal of Human Genetics*, 97, 111-124. <http://dx.doi.org/10.1016/j.ajhg.2015.05.020>
3. Wolf, N.I. & Smeitink, J.A. (2002). Mitochondrial disorders: a proposal for consensus diagnostic criteria in infants and children. *Neurology*, 59, 1402-5
4. Wortmann, S.B., Koolen, D.A., Smeitink, J.A., Heuvel, L. van den. & Rodenburg, R.J. (2015). Whole exome sequencing of suspected mitochondrial patients in clinical practice. *Journal of Inherited Metabolic Disease*, 38, 437-443. <http://dx.doi.org/10.1007/s10545-015-9823-y>
5. Calvo, S.E., Clauser, K.R. & Mootha, V.K. (2016). MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Research*, 44, D1251-D1257. <http://dx.doi.org/10.1093/nar/gkv1003>
6. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., ... Mesirov, J.P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 15545–15550. <http://dx.doi.org/10.1073/pnas.0506580102>
7. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., ... Mering, C. von. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, 43, D447-D452. <http://dx.doi.org/10.1093/nar/gku1003>
8. Szklarczyk, R., Megchelenbrink, W., Cizek, P., Ledent, M., Velemans, G., Szklarczyk, D. & Huynen, M.A. (2015). WeGET: predicting new genes for molecular systems by weighted co-expression. *Nucleic Acids Research*, 44, D567-D573. <http://dx.doi.org/10.1093/nar/gkv1228>
9. Bardella, C., Pollard, P.J. & Tomlinson, I. (2011). SDH mutations in cancer. *Biochimica et Biophysica Acta*, 1807, 1432-1443. <http://dx.doi.org/10.1016/j.bbabo.2011.07.003>
10. El-Hattab, A.W. and Scaglia, F. (2013) Mitochondrial DNA Depletion Syndromes: Review and Updates of Genetic Basis, Manifestations, and Therapeutic Options. *Neurotherapeutics*, 2, 186-198. <http://dx.doi.org/10.1007%2Fs13311-013-0177-6>
11. Kinghorn, K.J., Castillo-Quan, J.I., Bartolome, F., Angelova, P.R., Li, L., Pope, S., ... , Partridge, L., (2015) Loss of PLA2G6 leads to elevated mitochondrial lipid peroxidation and mitochondrial dysfunction. *Brain*, 7, 1801-16. <http://dx.doi.org/10.1093/brain/awv132>
12. Suzuki, T., Tanabe, K., Hara, I., Taniguchi, N. & Colavita, A. (2007). Dual enzymatic properties of the cytoplasmic peptide: N-glycanase in *C. elegans*. *Biochemical and Biophysical Research Communications*, 358, 837-841. <http://dx.doi.org/10.1016/j.bbrc.2007.04.199>
13. Robinson, P.N., Köhler, S., Oellrich, A., Wang, K., Mungall, C.J., Lewis, S.E., Smedley, D. (2014) Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Res.* 24(2): 340–348. <http://dx.doi.org/10.1101%2Fgr.160325.113>
14. Sifrim, A., Popovic, D., Tranchevent, L-C., Ardeshirdavani, A., Sakai, R., Konings, P., ... , Moreau, Y. (2013) eXtasy: variant prioritization by genomic data fusion. *Nat Methods.*, 11, 1083-4. <http://dx.doi.org/10.1038/nmeth.2656>
15. Javed, A., Agrawal, S., Ng, P.C. (2014) Phen-Gen: combining phenotype and genotype to analyze rare disorders. *Nature Methods* 9, 935-7. <http://dx.doi.org/10.1038/nmeth.3046>
16. Singleton, M.V., Guthery, S.L., Voelkerding, K.V., Chen, K., Kennedy, B., Margraf, R.L.,...,Yandell, M. (2014) Phevor combines multiple biomedical ontologies for accurate identification of disease-causing alleles in single individuals and small nuclear families. *Am J Hum Genet.*3;94(4):599-610. <http://dx.doi.org/10.1016/j.ajhg.2014.03.010>

APPENDICES

Appendix I: All query gene lists used in this report.

1: 'MitoAll'. Of the 1158 genes from Mitocarta, 351 were associated to phenotypes by HPO.

2: 'MitoSDHrem'. Same as list 1, but SDHx genes removed from list.

3: 'SuperMito'. Selection of list 2, genes that are involved in energy production of mitochondrion.

| Gene ID | Gene name | 1. | 2. | 3. |
|---------|-----------|----|----|----|
| 57505 | AARS2 | 1 | 1 | 1 |
| 10157 | AASS | 1 | 1 | 0 |
| 18 | ABAT | 1 | 1 | 0 |
| 10058 | ABCB6 | 1 | 1 | 0 |
| 22 | ABCB7 | 1 | 1 | 0 |
| 215 | ABCD1 | 1 | 1 | 0 |
| 5825 | ABCD3 | 1 | 1 | 0 |
| 31 | ACACA | 1 | 1 | 0 |
| 27034 | ACAD8 | 1 | 1 | 0 |
| 28976 | ACAD9 | 1 | 1 | 1 |
| 34 | ACADM | 1 | 1 | 0 |
| 35 | ACADS | 1 | 1 | 0 |
| 36 | ACADSB | 1 | 1 | 0 |
| 37 | ACADVL | 1 | 1 | 0 |
| 38 | ACAT1 | 1 | 1 | 0 |
| 50 | ACO2 | 1 | 1 | 1 |
| 51 | ACOX1 | 1 | 1 | 0 |
| 197322 | ACSF3 | 1 | 1 | 0 |
| 2182 | ACSL4 | 1 | 1 | 0 |
| 56997 | ADCK3 | 1 | 1 | 1 |
| 79934 | ADCK4 | 1 | 1 | 1 |
| 10939 | AFG3L2 | 1 | 1 | 0 |
| 55750 | AGK | 1 | 1 | 1 |
| 189 | AGXT | 1 | 1 | 0 |
| 9131 | AIFM1 | 1 | 1 | 1 |
| 204 | AK2 | 1 | 1 | 0 |
| 212 | ALAS2 | 1 | 1 | 0 |
| 5832 | ALDH18A1 | 1 | 1 | 0 |
| 217 | ALDH2 | 1 | 1 | 0 |
| 224 | ALDH3A2 | 1 | 1 | 0 |
| 8659 | ALDH4A1 | 1 | 1 | 0 |
| 7915 | ALDH5A1 | 1 | 1 | 0 |
| 4329 | ALDH6A1 | 1 | 1 | 0 |
| 501 | ALDH7A1 | 1 | 1 | 0 |
| 23600 | AMACR | 1 | 1 | 0 |
| 275 | AMT | 1 | 1 | 0 |
| 84334 | APOPT1 | 1 | 1 | 1 |
| 471 | ATIC | 1 | 1 | 0 |
| 498 | ATP5A1 | 1 | 1 | 1 |
| 514 | ATP5E | 1 | 1 | 1 |
| 91647 | ATPAF2 | 1 | 1 | 1 |
| 6311 | ATXN2 | 1 | 1 | 0 |
| 549 | AUH | 1 | 1 | 0 |
| 581 | BAX | 1 | 1 | 0 |
| 593 | BCKDHA | 1 | 1 | 0 |
| 594 | BCKDHB | 1 | 1 | 0 |
| 10295 | BCKDK | 1 | 1 | 0 |
| 596 | BCL2 | 1 | 1 | 0 |
| 617 | BCS1L | 1 | 1 | 1 |
| 388962 | BOLA3 | 1 | 1 | 1 |
| 56652 | C10orf2 | 1 | 1 | 1 |
| 91574 | C12orf65 | 1 | 1 | 1 |
| 763 | CA5A | 1 | 1 | 0 |
| 841 | CASP8 | 1 | 1 | 0 |

| Gene ID | Gene name | 1. | 2. | 3. |
|---------|-----------|----|----|----|
| 847 | CAT | 1 | 1 | 0 |
| 400916 | CHCHD10 | 1 | 1 | 1 |
| 493856 | CISD2 | 1 | 1 | 0 |
| 81570 | CLPB | 1 | 1 | 0 |
| 8192 | CLPP | 1 | 1 | 0 |
| 493753 | COA5 | 1 | 1 | 1 |
| 388753 | COA6 | 1 | 1 | 1 |
| 80347 | COASY | 1 | 1 | 0 |
| 1312 | COMT | 1 | 1 | 0 |
| 27235 | COQ2 | 1 | 1 | 1 |
| 51117 | COQ4 | 1 | 1 | 1 |
| 51004 | COQ6 | 1 | 1 | 1 |
| 57017 | COQ9 | 1 | 1 | 1 |
| 4512 | COX1 | 1 | 1 | 1 |
| 1352 | COX10 | 1 | 1 | 1 |
| 84987 | COX14 | 1 | 1 | 1 |
| 1355 | COX15 | 1 | 1 | 1 |
| 116228 | COX20 | 1 | 1 | 1 |
| 4514 | COX3 | 1 | 1 | 1 |
| 84701 | COX4I2 | 1 | 1 | 1 |
| 1337 | COX6A1 | 1 | 1 | 1 |
| 1340 | COX6B1 | 1 | 1 | 1 |
| 1349 | COX7B | 1 | 1 | 1 |
| 1371 | CPOX | 1 | 1 | 0 |
| 1373 | CPS1 | 1 | 1 | 0 |
| 1374 | CPT1A | 1 | 1 | 0 |
| 126129 | CPT1C | 1 | 1 | 0 |
| 1376 | CPT2 | 1 | 1 | 0 |
| 1528 | CYB5A | 1 | 1 | 0 |
| 1727 | CYB5R3 | 1 | 1 | 0 |
| 1537 | CYC1 | 1 | 1 | 1 |
| 54205 | CYCS | 1 | 1 | 1 |
| 1583 | CYP11A1 | 1 | 1 | 0 |
| 1585 | CYP11B2 | 1 | 1 | 0 |
| 1591 | CYP24A1 | 1 | 1 | 0 |
| 1593 | CYP27A1 | 1 | 1 | 0 |
| 1594 | CYP27B1 | 1 | 1 | 0 |
| 4519 | CYTB | 1 | 1 | 1 |
| 728294 | D2HGDH | 1 | 1 | 0 |
| 55157 | DARS2 | 1 | 1 | 1 |
| 1629 | DBT | 1 | 1 | 0 |
| 1666 | DECR1 | 1 | 1 | 0 |
| 1716 | DGUOK | 1 | 1 | 1 |
| 1718 | DHCR24 | 1 | 1 | 0 |
| 1723 | DHODH | 1 | 1 | 0 |
| 55526 | DHTKD1 | 1 | 1 | 0 |
| 56616 | DIABLO | 1 | 1 | 0 |
| 1737 | DLAT | 1 | 1 | 1 |
| 1738 | DLD | 1 | 1 | 1 |
| 29958 | DMGDH | 1 | 1 | 0 |
| 1760 | DMPK | 1 | 1 | 0 |
| 1763 | DNA2 | 1 | 1 | 1 |
| 131118 | DNAJC19 | 1 | 1 | 1 |
| 10059 | DNM1L | 1 | 1 | 1 |

| Gene ID | Gene name | 1. | 2. | 3. |
|---------|-----------|----|----|----|
| 124454 | EARS2 | 1 | 1 | 1 |
| 1892 | ECHS1 | 1 | 1 | 1 |
| 1962 | EHHADH | 1 | 1 | 0 |
| 60528 | ELAC2 | 1 | 1 | 1 |
| 2053 | EPHX2 | 1 | 1 | 0 |
| 2108 | ETFA | 1 | 1 | 0 |
| 2109 | ETFB | 1 | 1 | 0 |
| 2110 | ETFDH | 1 | 1 | 0 |
| 23474 | ETHE1 | 1 | 1 | 1 |
| 10667 | FARS2 | 1 | 1 | 1 |
| 22868 | FASTKD2 | 1 | 1 | 1 |
| 26235 | FBXL4 | 1 | 1 | 1 |
| 2235 | FECH | 1 | 1 | 0 |
| 2271 | FH | 1 | 1 | 1 |
| 60681 | FKBP10 | 1 | 1 | 0 |
| 55572 | FOXRED1 | 1 | 1 | 1 |
| 2495 | FTH1 | 1 | 1 | 0 |
| 2395 | FXN | 1 | 1 | 1 |
| 2617 | GARS | 1 | 1 | 1 |
| 2628 | GATM | 1 | 1 | 0 |
| 2639 | GCDH | 1 | 1 | 0 |
| 2653 | GCSH | 1 | 1 | 0 |
| 54332 | GDAP1 | 1 | 1 | 1 |
| 2671 | GFER | 1 | 1 | 1 |
| 85476 | GFM1 | 1 | 1 | 1 |
| 2710 | GK | 1 | 1 | 0 |
| 2731 | GLDC | 1 | 1 | 0 |
| 51218 | GLRX5 | 1 | 1 | 1 |
| 2746 | GLUD1 | 1 | 1 | 0 |
| 132158 | GLYCTK | 1 | 1 | 0 |
| 2821 | GPI | 1 | 1 | 0 |
| 84706 | GPT2 | 1 | 1 | 0 |
| 2876 | GPX1 | 1 | 1 | 0 |
| 9380 | GRHPR | 1 | 1 | 0 |
| 2936 | GSR | 1 | 1 | 0 |
| 84705 | GTPBP3 | 1 | 1 | 1 |
| 3033 | HADH | 1 | 1 | 0 |
| 3030 | HADHA | 1 | 1 | 0 |
| 3032 | HADHB | 1 | 1 | 0 |
| 23438 | HARS2 | 1 | 1 | 1 |
| 3052 | HCCS | 1 | 1 | 1 |
| 26275 | HIBCH | 1 | 1 | 0 |
| 3094 | HINT1 | 1 | 1 | 0 |
| 3098 | HK1 | 1 | 1 | 0 |
| 3145 | HMBS | 1 | 1 | 0 |
| 3155 | HMGCL | 1 | 1 | 0 |
| 3158 | HMGCS2 | 1 | 1 | 0 |
| 112817 | HOGA1 | 1 | 1 | 0 |
| 3028 | HSD17B10 | 1 | 1 | 1 |
| 3295 | HSD17B4 | 1 | 1 | 0 |
| 3329 | HSPD1 | 1 | 1 | 1 |
| 55699 | IARS2 | 1 | 1 | 1 |
| 200205 | IBA57 | 1 | 1 | 1 |
| 3418 | IDH2 | 1 | 1 | 1 |

| Gene ID | Gene name | 1. | 2. | 3. |
|---------|-----------|----|----|----|
| 3420 | IDH3B | 1 | 1 | 1 |
| 122961 | ISCA2 | 1 | 1 | 1 |
| 23479 | ISCU | 1 | 1 | 1 |
| 3712 | IVD | 1 | 1 | 0 |
| 3735 | KARS | 1 | 1 | 1 |
| 23095 | KIF1B | 1 | 1 | 0 |
| 3852 | KRT5 | 1 | 1 | 0 |
| 79944 | L2HGDH | 1 | 1 | 0 |
| 23395 | LARS2 | 1 | 1 | 1 |
| 11019 | LIAS | 1 | 1 | 1 |
| 51601 | LIPT1 | 1 | 1 | 1 |
| 9361 | LONP1 | 1 | 1 | 0 |
| 10128 | LRPPRC | 1 | 1 | 1 |
| 57128 | LYRM4 | 1 | 1 | 1 |
| 90624 | LYRM7 | 1 | 1 | 1 |
| 4128 | MAOA | 1 | 1 | 0 |
| 4129 | MAOB | 1 | 1 | 0 |
| 92935 | MARS2 | 1 | 1 | 1 |
| 56922 | MCCC1 | 1 | 1 | 0 |
| 64087 | MCCC2 | 1 | 1 | 0 |
| 84693 | MCEE | 1 | 1 | 0 |
| 9927 | MFN2 | 1 | 1 | 1 |
| 92667 | MGME1 | 1 | 1 | 1 |
| 10367 | MICU1 | 1 | 1 | 0 |
| 4292 | MLH1 | 1 | 1 | 0 |
| 23417 | MLYCD | 1 | 1 | 0 |
| 326625 | MMAB | 1 | 1 | 0 |
| 25974 | MMACHC | 1 | 1 | 0 |
| 27249 | MMADHC | 1 | 1 | 0 |
| 4337 | MOCS1 | 1 | 1 | 0 |
| 51660 | MPC1 | 1 | 1 | 1 |
| 4358 | MPV17 | 1 | 1 | 1 |
| 11222 | MRPL3 | 1 | 1 | 1 |
| 65080 | MRPL44 | 1 | 1 | 1 |
| 51021 | MRPS16 | 1 | 1 | 1 |
| 56945 | MRPS22 | 1 | 1 | 1 |
| 253827 | MSRB3 | 1 | 1 | 0 |
| 123263 | MTFMT | 1 | 1 | 1 |
| 25821 | MTO1 | 1 | 1 | 1 |
| 55149 | MTPAP | 1 | 1 | 1 |
| 4594 | MUT | 1 | 1 | 0 |
| 4595 | MUTYH | 1 | 1 | 0 |
| 133686 | NADK2 | 1 | 1 | 0 |
| 162417 | NAGS | 1 | 1 | 0 |
| 79731 | NARS2 | 1 | 1 | 1 |
| 4536 | ND2 | 1 | 1 | 1 |
| 4540 | ND5 | 1 | 1 | 1 |
| 4541 | ND6 | 1 | 1 | 1 |
| 4694 | NDUFA1 | 1 | 1 | 1 |
| 4705 | NDUFA10 | 1 | 1 | 1 |
| 126328 | NDUFA11 | 1 | 1 | 1 |
| 55967 | NDUFA12 | 1 | 1 | 1 |
| 4695 | NDUFA2 | 1 | 1 | 1 |
| 4704 | NDUFA9 | 1 | 1 | 1 |

| Gene ID | Gene name | 1. | 2. | 3. |
|---------|-----------|----|----|----|
| 51103 | NDUFAF1 | 1 | 1 | 1 |
| 91942 | NDUFAF2 | 1 | 1 | 1 |
| 25915 | NDUFAF3 | 1 | 1 | 1 |
| 29078 | NDUFAF4 | 1 | 1 | 1 |
| 79133 | NDUFAF5 | 1 | 1 | 1 |
| 137682 | NDUFAF6 | 1 | 1 | 1 |
| 54539 | NDUFB11 | 1 | 1 | 1 |
| 4709 | NDUFB3 | 1 | 1 | 1 |
| 4715 | NDUFB9 | 1 | 1 | 1 |
| 4719 | NDUFS1 | 1 | 1 | 1 |
| 4720 | NDUFS2 | 1 | 1 | 1 |
| 4722 | NDUFS3 | 1 | 1 | 1 |
| 4724 | NDUFS4 | 1 | 1 | 1 |
| 4726 | NDUFS6 | 1 | 1 | 1 |
| 374291 | NDUFS7 | 1 | 1 | 1 |
| 4728 | NDUFS8 | 1 | 1 | 1 |
| 4723 | NDUFV1 | 1 | 1 | 1 |
| 4729 | NDUFV2 | 1 | 1 | 1 |
| 27247 | NFU1 | 1 | 1 | 1 |
| 23530 | NNT | 1 | 1 | 0 |
| 4913 | NTHL1 | 1 | 1 | 0 |
| 80224 | NUBPL | 1 | 1 | 1 |
| 4942 | OAT | 1 | 1 | 0 |
| 4967 | OGDH | 1 | 1 | 1 |
| 4968 | OGG1 | 1 | 1 | 0 |
| 4976 | OPA1 | 1 | 1 | 1 |
| 80207 | OPA3 | 1 | 1 | 1 |
| 5009 | OTC | 1 | 1 | 0 |
| 5019 | OXCT1 | 1 | 1 | 0 |
| 5034 | P4HB | 1 | 1 | 0 |
| 51025 | PAM16 | 1 | 1 | 0 |
| 80025 | PANK2 | 1 | 1 | 0 |
| 11315 | PARK7 | 1 | 1 | 0 |
| 5091 | PC | 1 | 1 | 1 |
| 5095 | PCCA | 1 | 1 | 0 |
| 5096 | PCCB | 1 | 1 | 0 |
| 5106 | PCK2 | 1 | 1 | 0 |
| 5160 | PDHA1 | 1 | 1 | 1 |
| 5162 | PDHB | 1 | 1 | 1 |
| 8050 | PDHX | 1 | 1 | 1 |
| 5165 | PDK3 | 1 | 1 | 1 |
| 54704 | PDP1 | 1 | 1 | 1 |
| 23590 | PDSS1 | 1 | 1 | 1 |
| 57107 | PDSS2 | 1 | 1 | 1 |
| 100131 | PET100 | 1 | 1 | 1 |
| 801 | PET100 | 1 | 1 | 1 |
| 8799 | PEX11B | 1 | 1 | 0 |
| 5264 | PHYH | 1 | 1 | 0 |
| 65018 | PINK1 | 1 | 1 | 1 |
| 5313 | PKLR | 1 | 1 | 0 |
| 50640 | PNPLA8 | 1 | 1 | 1 |
| 55163 | PNPO | 1 | 1 | 0 |
| 87178 | PNPT1 | 1 | 1 | 1 |
| 5428 | POLG | 1 | 1 | 1 |

| Gene ID | Gene name | 1. | 2. | 3. |
|---------|-----------|----|----|----|
| 11232 | POLG2 | 1 | 1 | 1 |
| 152926 | PPM1K | 1 | 1 | 0 |
| 5498 | PPOX | 1 | 1 | 0 |
| 5625 | PRODH | 1 | 1 | 0 |
| 51651 | PTRH2 | 1 | 1 | 1 |
| 5805 | PTS | 1 | 1 | 0 |
| 80324 | PUS1 | 1 | 1 | 1 |
| 5831 | PYCR1 | 1 | 1 | 0 |
| 29920 | PYCR2 | 1 | 1 | 0 |
| 5860 | QDPR | 1 | 1 | 0 |
| 5917 | RARS | 1 | 1 | 0 |
| 57038 | RARS2 | 1 | 1 | 1 |
| 51109 | RDH11 | 1 | 1 | 0 |
| 9401 | RECQL4 | 1 | 1 | 1 |
| 55005 | RMND1 | 1 | 1 | 1 |
| 246243 | RNASEH1 | 1 | 1 | 1 |
| 22934 | RPIA | 1 | 1 | 0 |
| 6165 | RPL35A | 1 | 1 | 0 |
| 6208 | RPS14 | 1 | 1 | 0 |
| 54938 | SARS2 | 1 | 1 | 1 |
| 6341 | SCO1 | 1 | 1 | 1 |
| 9997 | SCO2 | 1 | 1 | 1 |
| 6342 | SCP2 | 1 | 1 | 0 |
| 6389 | SDHA | 1 | 0 | 0 |
| 644096 | SDHAF1 | 1 | 0 | 0 |
| 54949 | SDHAF2 | 1 | 0 | 0 |
| 6390 | SDHB | 1 | 0 | 0 |
| 6391 | SDHC | 1 | 0 | 0 |
| 6392 | SDHD | 1 | 0 | 0 |
| 79048 | SECISBP2 | 1 | 1 | 0 |
| 84947 | SERAC1 | 1 | 1 | 1 |
| 119559 | SFXN4 | 1 | 1 | 1 |
| 6566 | SLC16A1 | 1 | 1 | 0 |
| 6576 | SLC25A1 | 1 | 1 | 0 |
| 8604 | SLC25A12 | 1 | 1 | 0 |
| 10165 | SLC25A13 | 1 | 1 | 0 |
| 10166 | SLC25A15 | 1 | 1 | 0 |
| 60386 | SLC25A19 | 1 | 1 | 1 |
| 788 | SLC25A20 | 1 | 1 | 0 |
| 79751 | SLC25A22 | 1 | 1 | 0 |
| 5250 | SLC25A3 | 1 | 1 | 1 |
| 54977 | SLC25A38 | 1 | 1 | 0 |
| 291 | SLC25A4 | 1 | 1 | 1 |
| 91137 | SLC25A46 | 1 | 1 | 0 |
| 2542 | SLC37A4 | 1 | 1 | 0 |
| 9342 | SNAP29 | 1 | 1 | 0 |
| 6647 | SOD1 | 1 | 1 | 0 |
| 6687 | SPG7 | 1 | 1 | 0 |
| 6697 | SPR | 1 | 1 | 0 |
| 9517 | SPTLC2 | 1 | 1 | 0 |
| 6770 | STAR | 1 | 1 | 0 |
| 2040 | STOM | 1 | 1 | 0 |
| 8803 | SUCLA2 | 1 | 1 | 1 |
| 8802 | SUCLG1 | 1 | 1 | 1 |

| Gene ID | Gene name | 1. | 2. | 3. |
|---------|-----------|----|----|----|
| 79783 | SUGCT | 1 | 1 | 0 |
| 6821 | SUOX | 1 | 1 | 0 |
| 6834 | SURF1 | 1 | 1 | 1 |
| 51204 | TACO1 | 1 | 1 | 1 |
| 80222 | TARS2 | 1 | 1 | 1 |
| 10312 | TCIRG1 | 1 | 1 | 0 |
| 1678 | TIMM8A | 1 | 1 | 1 |
| 7084 | TK2 | 1 | 1 | 1 |
| 84233 | TMEM126A | 1 | 1 | 1 |
| 54968 | TMEM70 | 1 | 1 | 1 |
| 55217 | TMLHE | 1 | 1 | 0 |
| 7167 | TPI1 | 1 | 1 | 0 |
| 55687 | TRMU | 1 | 1 | 1 |
| 51095 | TRNT1 | 1 | 1 | 1 |
| 10102 | TSFM | 1 | 1 | 1 |
| 54902 | TTC19 | 1 | 1 | 1 |
| 10381 | TUBB3 | 1 | 1 | 0 |
| 7284 | TUFM | 1 | 1 | 1 |
| 10587 | TXNRD2 | 1 | 1 | 0 |
| 7374 | UNG | 1 | 1 | 0 |
| 84300 | UQCC2 | 1 | 1 | 1 |
| 7381 | UQCRB | 1 | 1 | 1 |
| 7385 | UQCRC2 | 1 | 1 | 1 |
| 27089 | UQCRQ | 1 | 1 | 1 |
| 57176 | VARA2 | 1 | 1 | 1 |
| 124997 | WDR81 | 1 | 1 | 0 |
| 63929 | XPBPEP3 | 1 | 1 | 0 |
| 51067 | YARS2 | 1 | 1 | 1 |

| rank | HPO ID | Phenotype description | N | q | nq | PS |
|------|-------------------|-----------------------------------------------------------------|----------|----------|------------|-------------|
| 1 | HP:0002490 | Increased CSF lactate | 49 | 49 | 0.1 | 11.98 |
| 2 | HP:0008316 | Abnormal mitochondria in muscle tissue | 21 | 21 | 0.1 | 10.76 |
| 3 | HP:0001404 | Hepatocellular necrosis | 18 | 18 | 0.1 | 10.54 |
| 4 | HP:0006965 | Acute necrotizing encephalopathy | 18 | 18 | 0.1 | 10.54 |
| 5 | HP:0012240 | Increased intramyocellular lipid droplets | 15 | 15 | 0.1 | 10.27 |
| 6 | HP:0003348 | Hyperalaninemia | 10 | 10 | 0.1 | 9.69 |
| 7 | HP:0006565 | Increased hepatocellular lipid droplets | 10 | 10 | 0.1 | 9.69 |
| 8 | HP:0001414 | Microvesicular hepatic steatosis | 8 | 8 | 0.1 | 9.37 |
| 9 | HP:0003150 | Glutaric aciduria | 8 | 8 | 0.1 | 9.37 |
| 10 | HP:0001112 | Leber optic atrophy | 6 | 6 | 0.1 | 8.95 |
| 11 | HP:0001117 | Sudden loss of visual acuity | 6 | 6 | 0.1 | 8.95 |
| 12 | HP:0001129 | Large central visual field defect | 6 | 6 | 0.1 | 8.95 |
| 13 | HP:0003219 | Ethylmalonic aciduria | 6 | 6 | 0.1 | 8.95 |
| 14 | HP:0008972 | Decreased activity of mitochondrial respiratory chain | 6 | 6 | 0.1 | 8.95 |
| 15 | HP:0001427 | Mitochondrial inheritance | 45 | 44 | 1 | 8.50 |
| 16 | HP:0000361 | Pulsatile tinnitus (tympanic paraganglioma) | 4 | 4 | 0.1 | 8.37 |
| 17 | HP:0002377 | Paraganglioma-related cranial nerve palsy | 4 | 4 | 0.1 | 8.37 |
| 18 | HP:0003001 | Glomus jugular tumor | 4 | 4 | 0.1 | 8.37 |
| 19 | HP:0030074 | Chemodectoma | 4 | 4 | 0.1 | 8.37 |
| 20 | HP:0003530 | Glutaric acidemia | 4 | 4 | 0.1 | 8.37 |
| 21 | HP:0006980 | Progressive leukoencephalopathy | 4 | 4 | 0.1 | 8.37 |
| 22 | HP:0008314 | Decreased activity of mitochondrial complex II | 4 | 4 | 0.1 | 8.37 |
| 23 | HP:0008344 | Elevated plasma branched chain amino acids | 4 | 4 | 0.1 | 8.37 |
| 24 | HP:0000740 | Anxiety (with pheochromocytoma) | 3 | 3 | 0.1 | 7.95 |
| 25 | HP:0001011 | Diaphoresis (with pheochromocytoma) | 3 | 3 | 0.1 | 7.95 |
| 26 | HP:0001606 | Vocal cord paralysis (caused by tumor impingement) | 3 | 3 | 0.1 | 7.95 |
| 27 | HP:0001613 | Hoarse voice (caused by tumor impingement) | 3 | 3 | 0.1 | 7.95 |
| 28 | HP:0001673 | Tachycardia (with pheochromocytoma) | 3 | 3 | 0.1 | 7.95 |
| 29 | HP:0001676 | Palpitations (with pheochromocytoma) | 3 | 3 | 0.1 | 7.95 |
| 30 | HP:0001686 | Loss of voice | 3 | 3 | 0.1 | 7.95 |
| 31 | HP:0002331 | Headache (with pheochromocytoma) | 3 | 3 | 0.1 | 7.95 |
| 32 | HP:0002640 | Hypertension associated with pheochromocytoma | 3 | 3 | 0.1 | 7.95 |
| 33 | HP:0004897 | Stress/infection-induced lactic acidosis | 3 | 3 | 0.1 | 7.95 |
| 34 | HP:0006737 | Extraadrenal pheochromocytoma | 3 | 3 | 0.1 | 7.95 |
| 35 | HP:0006748 | Adrenal pheochromocytoma | 3 | 3 | 0.1 | 7.95 |
| 36 | HP:0002161 | Hyperlysinemia | 3 | 3 | 0.1 | 7.95 |
| 37 | HP:0002614 | Hepatic periportal necrosis | 3 | 3 | 0.1 | 7.95 |
| 38 | HP:0002686 | Prenatal maternal abnormality | 3 | 3 | 0.1 | 7.95 |
| 39 | HP:0002928 | Decreased activity of the pyruvate dehydrogenase (PDH) complex | 3 | 3 | 0.1 | 7.95 |
| 40 | HP:0003287 | Abnormality of mitochondrial metabolism | 3 | 3 | 0.1 | 7.95 |
| 41 | HP:0003353 | Propionyl-CoA carboxylase deficiency | 3 | 3 | 0.1 | 7.95 |
| 42 | HP:0003490 | Defective dehydrogenation of isovaleryl CoA and butyryl CoA | 3 | 3 | 0.1 | 7.95 |
| 43 | HP:0003647 | Electron transfer flavoprotein-ubiquinone oxidoreductase defect | 3 | 3 | 0.1 | 7.95 |
| 44 | HP:0006799 | Basal ganglia cysts | 3 | 3 | 0.1 | 7.95 |
| 45 | HP:0011923 | Decreased activity of mitochondrial complex I | 3 | 3 | 0.1 | 7.95 |
| 46 | HP:0011924 | Decreased activity of mitochondrial complex III | 3 | 3 | 0.1 | 7.95 |
| 47 | HP:0002886 | Vagal paraganglioma | 2 | 2 | 0.1 | 7.37 |
| 48 | HP:0003334 | Elevated circulating catecholamine level | 2 | 2 | 0.1 | 7.37 |
| 49 | HP:0006715 | Glomus tympanicum paraganglioma | 2 | 2 | 0.1 | 7.37 |
| 50 | HP:0001958 | Nonketotic hypoglycemia | 2 | 2 | 0.1 | 7.37 |
| 51 | HP:0003344 | 3-Methylglutaric aciduria | 2 | 2 | 0.1 | 7.37 |
| 52 | HP:0003359 | Decreased urinary sulfate | 2 | 2 | 0.1 | 7.37 |
| 53 | HP:0003489 | Acute episodes of neuropathic symptoms | 2 | 2 | 0.1 | 7.37 |
| 54 | HP:0003572 | Low plasma citrulline | 2 | 2 | 0.1 | 7.37 |
| 55 | HP:0003643 | Sulfite oxidase deficiency | 2 | 2 | 0.1 | 7.37 |

Finding new candidate genes associated with molecular systems by computational screening for typical phenotypes | 2018

| | | | | | | |
|----|------------|---------------------------|---|---|-----|------|
| 56 | HP:0004448 | Fulminant hepatic failure | 2 | 2 | 0.1 | 7.37 |
| 57 | HP:0004925 | Chronic lactic acidosis | 2 | 2 | 0.1 | 7.37 |
| 58 | HP:0005974 | Episodic ketoacidosis | 2 | 2 | 0.1 | 7.37 |

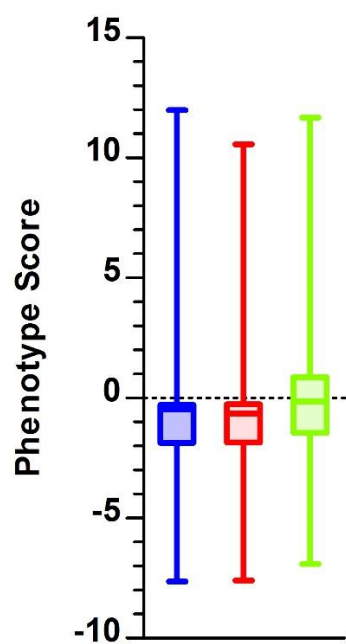
Appendix II. Best scoring phenotypes with their frequencies (that is N, q and nq) sorted on phenotype scores (PS). The phenotypes in grey rows are solely annotated by succinate

dehydrogenase genes, some of these phenotypes are not known to be a regular mitochondrial phenotype at all, but most likely originated from the pheochromocytoma / paraganglioma annotation of SDH-genes.

Appendix III Best scoring phenotypes of the SuperMito query. The light grey phenotypes ranked as 20-76 are not shown as they are all linked to solely one query gene, which results in a score of 7.52.

[illegible]

Appendix IV. Distribution of phenotype scores for the three different query lists: the MitoAll gene set (blue), the MitoSDHrem gene set (red) and the SuperMito gene set (green).



Appendix V. Enriched biological processes and KEGG pathways among the top 200 non-mitochondrial genes (using the SuperMito query).

KEGG pathways

| pathway ID | pathway description | No. | FDR | matching proteins |
|------------|-------------------------------------------------|-----|----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1100 | Metabolic pathways | 41 | 1.26e-10 | ACAT2, ADK, ALDOB, ALG11, ALG6, ALG9, AMPD2, ASPA, ASS1, B4GALNT1, DDOST, DHFR, DPYD, DPYS, ENO3, FBP1, FTCD, GAD1, GALC, GAMT, GCH1, GLUL, KYNU, MAT1A, NDST1, NT5C2, PCK1, PLA2G6, PLCE1, PNPLA2, POLR3A, PPT1, RRM2B, SLC33A1, ST3GAL5, STT3A, STT3B, SYNJ1, TH, TPK1, TYMP |
| 510 | N-Glycan biosynthesis | 7 | 5.59e-05 | ALG11, ALG6, ALG9, DDOST, RFT1, STT3A, STT3B |
| 4142 | Lysosome | 9 | 0.000257 | ARSA, CTSF, GALC, LAMP2, MFSD8, NPC1, NPC2, PPT1, PSAP |
| 4976 | Bile secretion | 7 | 0.000312 | ADCY5, ATP1A3, SLC10A2, SLC2A1, SLC4A4, SLC5A1, SLC9A1 |
| 970 | Aminoacyl-tRNA biosynthesis | 5 | 0.00339 | AARS, DARS, LARS, MARS, SEPSECS |
| 250 | Alanine, aspartate and glutamate metabolism | 4 | 0.0155 | ASPA, ASS1, GAD1, GLUL |
| 604 | Glycosphingolipid biosynthesis - ganglio series | 3 | 0.0155 | B4GALNT1, SLC33A1, ST3GAL5 |
| 240 | Pyrimidine metabolism | 6 | 0.0157 | DPYD, DPYS, NT5C2, POLR3A, RRM2B, TYMP |
| 4973 | Carbohydrate digestion and absorption | 4 | 0.0158 | ATP1A3, PIK3R5, SLC2A2, SLC5A1 |
| 1230 | Biosynthesis of amino acids | 5 | 0.0172 | ALDOB, ASS1, ENO3, GLUL, MAT1A |
| 3013 | RNA transport | 7 | 0.0183 | EIF2B1, EIF2B2, EIF2B3, EIF2B4, EIF2B5, NUP62, RANBP2 |
| 4964 | Proximal tubule bicarbonate reclamation | 3 | 0.0187 | ATP1A3, PCK1, SLC4A4 |
| 4919 | Thyroid hormone signaling pathway | 6 | 0.0203 | ATP1A3, MED17, PIK3R5, PLCE1, SLC2A1, SLC9A1 |
| 4141 | Protein processing in endoplasmic reticulum | 7 | 0.0218 | BCAP31, DDOST, DNAJB2, SAR1B, STT3A, STT3B, UBQLN2 |
| 4911 | Insulin secretion | 5 | 0.0223 | ADCY5, ATP1A3, SLC2A1, SLC2A2, SNAP25 |
| 4978 | Mineral absorption | 4 | 0.0235 | ATP1A3, CLCN2, FTL, SLC5A1 |
| 10 | Glycolysis / Gluconeogenesis | 4 | 0.0442 | ALDOB, ENO3, FBP1, PCK1 |
| 410 | beta-Alanine metabolism | 3 | 0.0442 | DPYD, DPYS, GAD1 |

Biological processes

| pathway ID | pathway description | No. | FDR | matching proteins |
|------------|------------------------------------------------|-----|----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| GO.0044281 | small molecule metabolic process | 58 | 4.59e-11 | AARS, ABHD12, ADCY5, AIMP1, ALDOB, ARSA, ASPA, ASS1, CA8, CHAT, CHKB, DARS, DNM2, DPYD, ENO3, ERLIN2, FA2H, FBP1, FOLR1, FTCD, GAD1, GALC, GLUL, GNB4, LARS, MARS, MAT1A, MMAA, MTRR, NDST1, NPC1, NPC2, NT5C2, NUP62, PCK1, PHKA1, PIK3R5, PLA2G6, PLCE1, PLP1, PNKP, PNPLA2, PSAP, RANBP2, SAR1B, SLC10A2, SLC19A3, SLC2A1, SLC2A2, SLC5A1, SLC9A1, SNAP25, SYNJ1, TH, TPK1, TTR, TYMP, VAPB |
| GO.0044710 | single-organism metabolic process | 81 | 9.57e-10 | AARS, ACAT2, ADCY5, AIMP1, ALDOB, ALG11, ALG6, ALG9, APTX, ARSA, ASPA, ASS1, CA8, CHAT, CHKB, CSF1R, CTSD, DARS, DDHD1, DDOST, DNAJB2, DNM2, DPYD, ENO3, ERLIN2, EXOSC3, FBP1, FBXO7, FOLR1, FTCD, GAD1, GALC, GLUL, GNB4, IGHMBP2, LARS, MARS, MAT1A, MMAA, MTRR, NPC1, NPC2, NT5C2, NUP62, PCK1, PCNA, PHKA1, PIK3R5, PLA2G6, PLCE1, PLP1, PNKP, PNPLA2, PPT1, PSAP, RANBP2, RFT1, SAR1B, SETX, SFTPB, SLC10A2, SLC19A3, SLC2A1, SLC2A2, SLC5A1, SLC5A7, SLC9A1, SNAP25, ST3GAL5, STT3A, STT3B, SYNJ1, TDP1, TH, TPK1, TYMP, UBQLN2, VAPB, VRK1, WDR45, ZFYVE26 |
| GO.0008366 | axon ensheathment | 13 | 3.22e-09 | ASPA, EGR2, EIF2B2, EIF2B4, EIF2B5, FA2H, JAM3, LAMA2, MTMR2, PMP22, PRX, SBF2, SH3TC2 |
| GO.1901564 | organonitrogen compound metabolic process | 46 | 3.22e-09 | AARS, ABHD12, ACY1, ADCY5, AIMP1, ALDOB, ARSA, ASPA, ASS1, ATP6AP2, CHAT, CHKB, DARS, DDOST, DPYD, EIF2B1, EIF2B2, EIF2B3, EIF2B4, EIF2B5, ENO3, FOLR1, FTCD, GAD1, GALC, GLUL, IGHMBP2, LARS, MARS, MAT1A, MMAA, MTRR, NDST1, NT5C2, PLA2G6, PPT1, PSAP, SEPSECS, SFTPB, SLC19A3, SLC9A1, ST3GAL5, TH, TPK1, TYMP, VAPB |
| GO.0042552 | myelination | 12 | 4e-08 | ASPA, EGR2, EIF2B2, EIF2B4, EIF2B5, FA2H, JAM3, LAMA2, MTMR2, PMP22, SBF2, SH3TC2 |
| GO.1901566 | organonitrogen compound biosynthetic process | 33 | 4.77e-08 | AARS, ADCY5, AIMP1, AMPD2, ASS1, B4GALNT1, CHAT, CHKB, DARS, DDOST, DHFR, DPYD, EIF2B1, EIF2B2, EIF2B3, EIF2B4, EIF2B5, GAMT, GLUL, IGHMBP2, KYNU, LARS, MARS, MAT1A, MMAA, MTRR, NDST1, SEPSECS, ST3GAL5, TH, TPK1, TYMP, VAPB |
| GO.0044712 | single-organism catabolic process | 31 | 1.27e-07 | ALDOB, ASL, ASPA, CTSD, DDHD1, DNAJB2, DPYD, DPYS, ENO3, ERLIN2, FBXO7, FTCD, GAD1, GALC, GLUL, LDHA, MMAA, MTMR2, MTRR, NT5C2, PHKA1, PLA2G6, PLCE1, PNPLA2, PPT1, SLC9A1, STT3B, SYNJ1, TYMP, UBQLN2, WDR45 |
| GO.0006520 | cellular amino acid metabolic process | 19 | 4.08e-07 | AARS, ABHD12, ACY1, AIMP1, ASPA, ASS1, DARS, DPYD, DPYS, FOLR1, FTCD, GAD1, GCH1, GLUL, LARS, MARS, MAT1A, MTRR, TH |
| GO.0043648 | dicarboxylic acid metabolic process | 11 | 5.17e-07 | ASPA, ASS1, DHFR, FOLR1, FTCD, GAD1, GLUL, KYNU, MTRR, PCK1, TH |
| GO.0019752 | carboxylic acid metabolic process | 28 | 8.71e-07 | AARS, ABHD12, ACY1, AIMP1, ALDOB, ASPA, ASS1, DARS, DPYD, DPYS, ENO3, FA2H, FOLR1, FTCD, GAD1, GCH1, GLUL, LARS, MARS, MAT1A, MMAA, MTRR, NPC1, PCK1, PLP1, SLC10A2, SLC2A1, TH |
| GO.0014003 | oligodendrocyte development | 8 | 1.17e-06 | ASPA, EIF2B1, EIF2B2, EIF2B3, EIF2B4, EIF2B5, FA2H, PLP1 |
| GO.0043603 | cellular amide metabolic process | 24 | 1.59e-06 | AARS, AIMP1, ASL, ASS1, ATP6AP2, B4GALNT1, DARS, DDOST, DHFR, EIF2B1, EIF2B2, EIF2B3, EIF2B4, EIF2B5, FOLR1, FTCD, GALC, GCH1, IGHMBP2, LARS, MARS, MTRR, SEPSECS, ST3GAL5 |
| GO.0021782 | glial cell development | 10 | 1.64e-06 | ASPA, EIF2B1, EIF2B2, EIF2B3, EIF2B4, EIF2B5, FA2H, LAMA2, PLP1, SH3TC2 |
| GO.1901135 | carbohydrate derivative metabolic process | 31 | 1.7e-06 | ABHD12, ADCY5, ALDOB, ALG11, ALG13, ALG6, ALG9, ARSA, DDOST, DPYD, DPYS, ENO3, FBP1, GALC, GAMT, LDHA, MAT1A, MTRR, NDST1, NPC1, NT5C2, PSAP, RFT1, RRM2B, SAR1B, SLC9A1, ST3GAL5, STT3A, STT3B, TH, TYMP |
| GO.0043436 | oxoacid metabolic process | 29 | 3.27e-06 | AARS, ABHD12, ACY1, AIMP1, ALDOB, ASPA, ASS1, DARS, DPYD, DPYS, ENO3, FA2H, FOLR1, FTCD, GAD1, GCH1, GLUL, LARS, MARS, MAT1A, MMAA, MTRR, NDST1, NPC1, PCK1, PLP1, SLC10A2, SLC2A1, TH |
| GO.0044711 | single-organism biosynthetic process | 35 | 4.41e-06 | ADCY5, ALDOB, AMPD2, ASS1, B4GALNT1, CHAT, CHKB, DHFR, DPYD, ENO3, FA2H, FBP1, GAMT, GCH1, GLUL, KYNU, MAT1A, MMAA, MTMR2, MTRR, PCK1, PCNA, PIK3R5, PLA2G6, PLCE1, PLP1, PNPLA2, RRM2B, SLC2A1, SLC5A7, ST3GAL5, SYNJ1, TH, TYMP, VAPB |
| GO.0005975 | carbohydrate metabolic process | 26 | 8.2e-06 | AIMP1, ALDOB, ALG11, ALG6, ALG9, DDOST, ENO3, FBP1, GALC, LDHA, MTMR2, NPC1, NUP62, PCK1, PHKA1, RANBP2, RFT1, SAR1B, SLC2A1, SLC2A2, SLC5A1, SLC9A1, ST3GAL5, STT3A, STT3B, SYNJ1 |
| GO.0043604 | amide biosynthetic process | 19 | 1.12e-05 | AARS, AIMP1, ASL, ASS1, B4GALNT1, DARS, DDOST, DHFR, EIF2B1, EIF2B2, EIF2B3, EIF2B4, EIF2B5, GCH1, IGHMBP2, LARS, MARS, SEPSECS, ST3GAL5 |
| GO.0044723 | single-organism carbohydrate metabolic process | 22 | 1.72e-05 | AIMP1, ALDOB, ALG11, ALG6, ALG9, B4GALNT1, DDOST, ENO3, FBP1, LDHA, MTMR2, NDST1, NPC1, PCK1, PHKA1, RFT1, SAR1B, SLC2A1, ST3GAL5, STT3A, STT3B, SYNJ1 |
| GO.0044765 | single-organism transport | 54 | 1.72e-05 | ABCA3, ADCY5, ATP1A3, BCAP31, CASQ1, CHAT, CHRNE, CLCN2, DDOST, DNAJC6, DNM2, EGR2, FBXO7, FOLR1, GAD1, GJB1, GJC2, GLUL, GRID2, GRIK2, KCNC1, KCNT1, KIAA0226, KIF1C, KIF5A, LAMP2, LRSAM1, MFSD8, NPC1, NPC2, PFN1, PLA2G6, PRSS12, PSAP, RAB7A, RANBP2, RFT1, SAR1B, SLC10A2, SLC19A3, SLC22A5, SLC33A1, SLC4A4, SLC52A1, SLC5A7, SLC9A1, SNAP25, SYNJ1, SYT2, TH, TTR, UCHL1, VAPB, VPS53 |

Appendix

Appendix VI: Genes that are included in the list of mutations found in multiple patients with secondary mitochondrial disorders. The WeGET results are shown in the second and third column: the p-value of the genes coexpressed with the SuperMito gene list and the corresponding rank in the total gene list that WeGET provides as output. The third column contains the gene scores compiled from the SuperMito query. Bold genes have significant co-expression ($P < 0.05$).

| | Gene | P-value | rank | GS |
|----|---------------|--------------|-------------|--------------|
| 1 | PDIA6 | 0.294 | 5456 | NA |
| 2 | GRIN3B | 0.671 | 10589 | NA |
| 3 | PLEK | 1.000 | 18382 | NA |
| 4 | TBR1 | 1.000 | 20202 | NA |
| 5 | PLA2G6 | 0.594 | 9498 | 26.45 |
| 6 | FA2H | 0.926 | 14763 | 23.10 |
| 7 | LARS | 0.058 | 1891 | 15.39 |
| 8 | GALC | 0.539 | 8720 | 14.61 |
| 9 | RANBP2 | 0.312 | 5688 | 11.69 |
| 10 | EIF2B3 | 0.004 | 453 | 9.25 |
| 11 | ALG13 | 0.162 | 3672 | 7.50 |
| 12 | CHRNE | 1.000 | 19029 | 7.38 |
| 13 | ALG11 | NA | NA | 5.34 |
| 14 | LAMA2 | 0.693 | 10898 | 5.03 |
| 15 | DNAJC3 | 0.545 | 8779 | 3.92 |
| 16 | NGLY1 | 0.022 | 1056 | 3.85 |
| 17 | CLN8 | 0.590 | 9448 | 2.95 |
| 18 | MYF6 | 0.766 | 11987 | 2.20 |
| 19 | HCN1 | 1.000 | 17902 | 0.31 |
| 20 | MIP | 1.000 | 19566 | -2.51 |
| 21 | UPB1 | 0.817 | 12816 | -2.77 |
| 22 | CC2D1A | 0.712 | 11192 | -3.67 |
| 23 | TTN | 0.301 | 5548 | -3.77 |
| 24 | TFR2 | 0.577 | 9272 | -4.81 |
| 25 | ALG14 | 0.033 | 1340 | -5.25 |
| 26 | CLN6 | 0.666 | 10497 | -5.51 |
| 27 | CEL | 1.000 | 17740 | -5.71 |
| 28 | PPP2R5D | 0.173 | 3812 | -5.89 |
| 29 | EGFR | 1.000 | 17247 | -7.30 |
| 30 | ACTA1 | 0.376 | 6567 | -7.86 |
| 31 | SCN1A | 0.978 | 15799 | -7.90 |
| 32 | SEPN1 | 0.784 | 12265 | -9.17 |
| 33 | MYBPC3 | 0.090 | 2504 | -9.85 |
| 34 | STXBP1 | 1.000 | 16764 | -10.63 |
| 35 | SAMHD1 | 1.000 | 16500 | -10.97 |
| 36 | ASPM | 0.726 | 11409 | -11.30 |
| 37 | KCNQ2 | 0.671 | 10608 | -11.60 |
| 38 | TCN2 | 0.727 | 11423 | -14.01 |
| 39 | HSD3B7 | 0.259 | 5011 | -14.22 |
| 40 | MAGT1 | 0.176 | 3847 | -14.68 |
| 41 | SLC26A3 | 0.736 | 11530 | -18.81 |
| 42 | SLC3A1 | 0.431 | 7283 | -24.91 |
| 43 | CTNNB1 | 0.376 | 6557 | -26.21 |
| 44 | ANO5 | 0.158 | 3604 | -26.95 |
| 45 | TMEM173 | 1.000 | 16284 | -37.94 |
| 46 | HEXB | 0.139 | 3307 | -46.38 |
| 47 | KIAA058 | 0.778 | 12162 | -49.44 |
| 48 | EPG5 | 0.478 | 7901 | -62.30 |
| 49 | IFIH1 | 1.000 | 16811 | -65.83 |
| 50 | COL5A1 | 1.000 | 19239 | -103.38 |
| 51 | ARID1B | 1.000 | 16345 | -109.81 |
| 52 | COL4A1 | 0.401 | 6883 | -111.00 |
| 53 | KDM6A | 1.000 | 17848 | -135.90 |
| 54 | NOTCH3 | 0.807 | 12645 | -223.70 |
| 55 | SETBP1 | 0.576 | 9262 | -245.05 |